

1

COMPARABLE DISPARITY IN THE APPENDICULAR SKELETON ACROSS THE FISH-
TETRAPOD TRANSITION, AND THE MORPHOLOGICAL GAP BETWEEN FISH AND
TETRAPOD POSTCRANIA

by MARCELLO RUTA ¹ and MATTHEW A. WILLS ²

¹ School of Life Sciences, University of Lincoln, Green Lane, Lincoln LN6 7DL, UK; e-mail:

mruta@lincoln.ac.uk

² Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY,

UK; e-mail: m.a.wills@bath.ac.uk

Abstract: Appendicular skeletal traits are used to quantify changes in morphological disparity and morphospace occupation across the fish-tetrapod transition, and to explore the informativeness of different data partitions in phylogeny reconstruction. Anterior appendicular data yield trees that differ little from those built from the full character set, whilst posterior appendicular data result in considerable loss of phylogenetic resolution and tree branch rearrangements. Overall, there is significant incongruence in the signals associated with pectoral and pelvic data. The appendicular skeletons of fish and tetrapods attain similar levels of morphological disparity (at least when data are rarefied at the maximum sample size for fish in our study) and occupy similarly sized regions of morphospace. However, fish appear more dispersed in morphospace than tetrapods do. All taxa show a heterogeneous distribution in morphospace, and there is a clear separation between fish and tetrapods despite the presence of several evolutionarily intermediate taxa.

Key words: empirical morphospace, fins, fish, girdles, limbs, tetrapods.

THE origin of limbed vertebrates (tetrapods) from fish is among the most iconic, best known, and best studied of all major evolutionary transitions (Coates *et al.* 2008; Shubin, 2009; Clack 2009, 2012). This event entailed remarkable changes in structural complexity, ecological variety, and morphofunctional diversity (Coates *et al.* 2007; Shubin *et al.* 2009; Pierce *et al.* 2013; Neenan *et al.* 2014), and is therefore of considerable significance to biologists interested in tempo and mode of evolutionary radiations. Phylogenetic, palaeoecological, and functional aspects of the fish-tetrapod transition are most appropriately addressed by analysing changes in the appendicular skeleton. This is because the elaborate architecture of fins, limbs, and girdles offers a rich source of cladistic and biomechanical data (e.g. Clack 2012; Pierce *et al.* 2012; see also Ruta 2011), and permits detailed investigations into patterns and rates of trait change near the origin of a major animal radiation (e.g. Clack 2009, 2012; Coates *et al.* 2002, 2008; Ruta *et al.* 2006; Wagner *et al.* 2006). More broadly, fossil, embryological, and comparative anatomical data on the appendicular skeleton have promoted the interdisciplinary dialogue between evolutionary and developmental biologists (e.g. Coates *et al.* 2007; Hall 2007; Shubin *et al.* 2009; Clack 2012).

The last ten years have witnessed an astonishing proliferation of fossil finds spanning the fish-tetrapod transition, spurring the publication of increasingly detailed and refined matrices of skeletal characters and significant novel interpretations of existing data (e.g. Boisvert *et al.* 2008; Callier *et al.* 2009; Ahlberg 2011; Pierce *et al.* 2012, 2013; Smithson and Clack 2013; Anderson *et al.* 2015). However, recent cladistic analyses vary in the extent to which appendicular characters are atomised and coded. In his compendium of appendicular characters for some of the most adequately known fin- and limb-bearing stem-group tetrapods and early-branching crown-group tetrapod clades, Ruta (2011) sought to establish whether appendicular data alone are able to retrieve taxon relationships that are congruent with those inferred from more comprehensive data matrices. As part of ongoing research on models of evolutionary transformation during major adaptive radiations, we employ herewith the data in Ruta (2011) to investigate changes in the structural complexity of fins, limbs, and girdles across the fish-tetrapod transition, and to scrutinize further the issue of phylogenetic

informativeness in appendicular traits. We address two hypotheses. The first is that the appendicular skeletons of the fish-like relatives of tetrapods were less morphologically disparate than those of the digit-bearing taxa. The second is that characters of the pelvic and pectoral girdles support different phylogenies and define empirical morphospaces with significantly different characteristics. In some scenarios of adaptive diversification, our first hypothesis implies that morphological complexity increased coincident with, or immediately after, the onset of ecological innovation (specifically, the water-land transition) (e.g. Benton 2015; Erwin 2015). Our second hypothesis is that anterior and posterior appendicular characters did not evolve in tandem, as expounded by Coates *et al.* (2002).

The present paper adds to the growing number of analyses of morphological disparity in fossil groups (e.g. Briggs *et al.* 1992; Wills *et al.* 1994; Wills 1998*a, b*; Lofgren *et al.* 2003; Ruta *et al.* 2006; Adamowicz *et al.* 2008; Brusatte *et al.* 2008*a, b*, 2011; Ruta 2009; Cisneros and Ruta 2010; Young *et al.* 2010; Prentice *et al.* 2011; Thorne *et al.* 2011; Ausich and Deline 2012; Bapst *et al.* 2012; Benson *et al.* 2012; Brusatte *et al.* 2012; Foth *et al.*, 2012; Wills *et al.* 2012; Gerber 2013; Hughes *et al.* 2013; Ruta *et al.* 2013*a, b*; Stubbs *et al.* 2013; Smith *et al.* 2014; Trap 2014; Xue *et al.* 2015). Unlike most previous contributions, however, we focus on a single morphofunctional complex (the appendicular skeleton). In this, our approach resembles traditional and geometric morphometric analyses where specifically delimited constellations of landmarks or measurements of individual structures are used (e.g. Zelditch *et al.* 2004; for other examples of disparity studies based on a set of discrete traits, see Meloro and Jones 2012 and Grossnickle and Polly 2013). Therefore, results should be interpreted exclusively in the framework of the selected data (e.g. Wills 2001). Thus, our target is a well-studied and character-rich skeletal system that is analysed in a set of taxa bracketing narrowly a specific internode on the tetrapod stem-group, where key changes of interest took place. For species selection criteria, see Ruta *et al.* (2003), Ruta and Coates (2007), and Ruta (2011). ~~In the~~ Appendix [S2 \(Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.1hr49/2>\)](#), ~~we~~ reports the data matrix as a NEXUS file for PAUP* (Swofford 2002). For the complete character list, see Ruta (2011).

AIMS

This paper has ~~two~~three major aims. (1) We discuss the results of character partition experiments (Figs 1, 2) based on the data matrix of Ruta (2011). Such experiments are designed to investigate whether different data sets (i.e. anterior vs. posterior appendicular characters) carry comparable levels of phylogenetic signal (and, as a corollary, if a weak but correct signal shapes the tree exclusively when the partitions are combined), and whether they impact differently on disparity profiles at the fish-tetrapod transition. (2) ~~We use the data matrix of Ruta (2011) to~~ quantify changes in morphological disparity during the emergence of limbed vertebrates (Figs 3, 4). (23) We construct an empirical morphospace in which we quantify patterns of dispersal and clustering of species (Figs 5–8), comparing their observed distribution to a null model of complete spatial randomness (for statistical and analytical procedures, see Baddeley and Turner 2005; Hammer and Harper 2006). Finally, we discuss briefly the distribution of unique states in the Ruta (2011) matrix (Fig. 9). ~~We also discuss the results of character partition experiments. With these experiments, we ask whether different sets of data (i.e. anterior vs. posterior appendicular characters) carry comparable levels of phylogenetic signal (and, as a corollary, if a weak but correct signal shapes the tree exclusively when the partitions are combined), and whether they impact differently on disparity profiles at the fish-tetrapod transition.~~

MATERIAL AND METHODS

Disparity analyses

Group delimitations. Our taxonomic sample is divided into fin- and limb-bearing species. This subdivision offers a useful framework for discussing disparity changes in relation to the origin of an

easily identifiable novelty – the digits (e.g. Coates *et al.* 2002; Shubin *et al.* 2006; Hall 2007; Johanson *et al.* 2007; Boisvert *et al.* 2008; Woltering and Duboule 2010; Zhang *et al.* 2010; Ahlberg 2011). The ‘boundary’ between fin- and limb-bearing taxa was placed on the internode between elpistostegid-grade members of the tetrapod stem-group (i.e. *Panderichthys*; *Tiktaalik*) and all more crownward taxa (Text-fig. 1A). For some taxa, assignment to one or the other of these two categories took into account the most recent comprehensive phylogenies. A remarkable example is represented by the Catskill humerus (e.g. Shubin *et al.* 2004), attributed to digit-bearing taxa ~~here-in~~ the present paper specifically in order to bias our analyses towards ~~one-the first~~ of our working hypotheses (see above), that the appendicular skeleton of fin-bearing taxa is less disparate than that of limb-bearing taxa. At present, we do not know whether the Catskill humerus belonged to a limbed or a finned animal. By placing it among digit-bearing taxa, we enforce our null hypothesis that the range of appendicular morphologies – and hence complexity – of those taxa was greater than that of fish-like tetrapodomorphs. This way, we impose a more stringent scenario for the outcome of statistical tests of morphological differences between the two taxon sets. If in fact the morphological disparity of the finned taxa were found to be comparable to, or higher than, that of the limbed taxa, then we would place greater confidence in the performance of the tests in rejecting the null hypothesis.

Autapomorphies. We did not add autapomorphies to Ruta’s (2011) matrix for numerous pragmatic reasons, which we detail elsewhere (Ruta, Benton and Wills in preparation). Cisneros and Ruta (2010) provided a preliminary discussion of the rationale behind the exclusion of autapomorphies. Briefly, the practice of expressly enumerating autapomorphies in a data matrix is somewhat more arbitrary and subjective than the coding of synapomorphies. Unless the autapomorphies are highly concentrated in a small number of terminal taxa, however, their impact upon the distribution of taxa in morphospace is likely to be negligible, at least when either a variance-covariance matrix between characters (as in Principal Component Analysis) or a matrix of pairwise Euclidean (or generalised

Euclidean) distances between taxa (as in Principal Coordinates Analysis) are employed (e.g. Wills 1998a). Specifically, taxa are distributed very slightly more centrifugally when autapomorphies are included than when they are excluded. Where the distribution of autapomorphic states across the taxonomic exemplars is indistinguishable from random, the impact of including or excluding them is even less likely to be significant. Ruta (2011) attempted to accommodate as much variation in the morphology of the appendicular skeleton as possible (for similar protocols, see Ruta and Bolt 2008; Ruta 2009) and although he did not code for autapomorphies, instances of unique states within multistate characters were represented in his matrix. Here, we investigate the distribution of those unique states as a proxy for the distribution of the autapomorphies overall. The actual distribution of unique states was tested against a null best-fit ~~P~~poisson distribution for the data using chi-squared and maximum likelihood optimisations. The deviation of the actual distribution from this null was assessed using the VCD v. 1.4-0 package (Meyer *et al.* 2015) in the 'R' (v. 3.1.2) environment for statistical computing and graphics (<http://www.R-project.org/>).

Analytical protocols. We offer only a summary of the methods used for disparity analyses, as these are discussed extensively by Wills *et al.* (1994), Wills (1998a, b), Brusatte *et al.* (2008a, b, 2011), Ruta (2009), Cisneros and Ruta (2010), Young *et al.* (2010), Prentice *et al.* (2011), Thorne *et al.* (2011), Hughes *et al.* (2013), and Ruta *et al.* (2013b). With all characters unordered and of unit weight, we first converted the data matrix of Ruta (2011) into a dissimilarity matrix of intertaxon generalised Euclidean distances, employing MATRIX v. 1.0 (e.g. Wills *et al.* 1994; Wills 1998a).

~~This~~ The distance matrix (Appendix S1; *Dryad Digital Repository*: <http://dx.doi.org/10.5061/dryad.1hr49/1>) was then subjected to a Principal Coordinates Analysis (PCoA) in GINKGO v. 1.5.5 (Cáceres *et al.* 2007), with Cailliez correction for negative eigenvalues (Cailliez 1983). For disparity calculations (see below), we included sufficient successive PCo axes to explain 90 percent of the total variance in the data. The scores (i.e. the PCo coordinates) of taxa on those axes (Appendix S1; *Dryad Digital Repository*: <http://dx.doi.org/10.5061/dryad.1hr49/1>)

were used to calculate disparity indices in RARE v. 1.2 (e.g. Wills *et al.* 1994; Wills 1998a). All analyses and calculations were performed three times: once for the complete matrix and once each for the anterior and posterior appendicular character partitions. Four disparity indices were used: sums and products of ranges and variances (with all products being scaled to a single dimension). As well as calculating indices for all fin-bearing and digit-bearing (or presumed digit-bearing) taxa, we also used rarefaction analysis within these two groups in order to quantify the sensitivity of our results to perturbations in taxon sample and to sample size differences. This is because range-based metrics are affected chiefly by sample size (less so by the splitting or lumping of taxa), whereas variance-based metrics are influenced for the most part by taxon subdivision and merging (less so by sample size, at least when sampling is random; see Van Valen 1974; Wills *et al.* 1994; Ciampaglio *et al.* 2001; Villier and Korn 2004). The mean and median disparity values with approximate 90 percent confidence intervals were estimated using 10,000 bootstrap replicates (Figs 3, 4).

Two-dimensional plots of taxa using combinations of the first three PCo axes are illustrated in Figure 5A–C. Also, we explored in brief the results obtained from a Non-Metric Multidimensional Scaling (NMDS) analysis applied to the generalised intertaxon Euclidean distances derived from the entire character matrix. The taxon distribution in the plane delimited by the first two NMDS axes is shown in Figure 5D. For brevity, we have not conducted NMDS analyses on character partitions.

~~our~~ Our choice of the two multivariate methods (PCoA; NMDS) is underpinned by a purely exploratory approach. In practice, most of our discussion is based upon PCoA. For a detailed treatment of these methods, see also Legendre and Legendre (1998) and Hammer and Harper (2006).

Statistical protocols. The PCo scores were used in three permutation tests to quantify differences in the manner in which fish and tetrapods occupied their respective regions of morphospace. First, we tested if the variances of scores in these two groups were different using a permutation-based, non-

parametric multivariate analysis of variance (hereafter, npMANOVA) (Anderson 2001; Hammer and Harper 2006). Second, we tested if the ranked distances between taxa within their actual groups (i.e. fish or tetrapods) were different from the ranked distances between taxa assigned to the two groups at random. Formally, the null hypothesis of this analysis of similarity (hereafter, ANOSIM) was that ranked dissimilarities within groups have equal median and range (Clarke 1993; Hammer and Harper 2006). The third test – two-group permutation (hereafter, TGP) – evaluates the equality of the means of two groups via a permutation procedure. This was assessed by randomly assigning taxa to the two groups several times, each time calculating the new mean values (Hammer and Harper 2006). The groups' means were compared to the distribution of the mean values thus obtained. All tests were run in PAST v. 2.12 (Hammer *et al.* 2001; Hammer and Harper 2006) using 1000 permutations in all cases. Each test was applied to the entire dataset, and to each of the two character partitions (anterior and posterior appendicular characters). For all three tests, we also explored the effects of including a variable number of multivariate axes. Specifically, variants of npMANOVA, ANOSIM, and TGP used the first two, first three, and first five successive axes, as well as multiples of five axes up to, and including, the number of axes needed to encompass at least 90 percent of the total variance. For example, 36 axes account for 90 percent of the total variance when the full character matrix is employed. Therefore, we repeated npMANOVA, ANOSIM, and TGP tests with 2, 3, 5, 10, 15, 20, 25, 30, 35, and 36 axes (~~Table-Table S1~~: *Dryad Digital Repository*: <http://dx.doi.org/10.5061/dryad.1hr49/3>).

Using GINKGO, we produced tabulations of the pairwise PCo distances between taxa using the whole character set and the two appendicular character partitions. We then evaluated the degree of congruence among the three PCo distance matrices as well as their degree of similarity using the software CADM (Legendre and Lapointe 2004). CADM implements a global test of congruence using an extension of Mantel's (1967) test for the correlation between two distance matrices, and operates by randomly permuting the matrices independent of each other. CADM further performs *a posteriori* tests of the congruence of each individual matrix (a single matrix at a time is permuted) relative to

Formatted: Font: Italic

Formatted: Font: Times New Roman, 12 pt

all the others (the null hypothesis is that an individual matrix is incongruent with each of the others in turn). For analyses in CADM, which is now part of the APE (v. 3.0-8) package in 'R' (Paradis *et al.* 2004), we carried out 999 permutations of matrix structure, each matrix having equal unit weight.

In addition to the CADM analyses, we performed tests of concordance of taxon distributions in the multivariate morphospace plots derived from the whole character set and from each of the two character partition sets, using the Procrusean superimposition method devised by Peres-Neto and Jackson (2001; see Hetherington *et al.* 2015 for a recent application), and implemented in the VEGAN (v. 3.0-8) package in 'R' (Oksanen *et al.* 2013). In all pairwise comparisons among the three PCo distance matrices, the Procrusean superimposition method rotates a matrix such that it achieves maximum possible similarity with another matrix after the sum of squared differences between the same data points (taxa) in the two matrices is minimised. We used 999 permutations of matrix structure to evaluate the significance of the point distributions between any two matrices.

Finally, we tested whether five taxa that bracket narrowly the fish-tetrapod boundary in Figure 1A (*Panderichthys*, *Tiktaalik*, the Catskill humerus, *Acanthostega*, and *Ichthyostega*) were closer to other fin-bearing or digit-bearing taxa. For each of these 'boundary' taxa in turn, we first tagged the PCoA distances to all other 'non-boundary' taxa as either 'fish' or 'tetrapod', and then subjected the distances thus grouped to a Mann-Whitney test in PAST (null hypothesis: equality of medians in two independent samples; e.g. Hammer and Harper 2006).

Spatial analyses

Analytical protocols. In order to characterize patterns of taxon distribution in morphospace, we used two-dimensional point pattern analysis in the SPATSTAT v. 1.21-2 package (Baddeley and Turner 2005) in 'R'. We utilized the scores on the first two PCo axes for the analysis involving the full character matrix. In addition, we analyzed the taxon distribution using coordinates on the two NMDS dimensions. For all bivariate scatterplots, we first estimated the spatial bounds of the point

distribution with the Ripley-Rasson method (see Ripley and Rasson 1977). This generated a two-dimensional rectangular plane beyond the outermost data points, and representing an optimal observation window. We then produced a kernel density estimate of the probability distributions of taxa within this window. This is illustrated using bivariate density plots (e.g. Lucy *et al.* 2002; Fig. 6A, B) as well as perspective plots (Fig. 6C, D; e.g. Wand and Jones 1995) (see also Wills *et al.* 2012 for an example). In the bivariate density plots, different colour hues represent areas with identical probability density values. In the perspective plots, the ‘peaks’ and ‘troughs’ represent, respectively, higher and lower probability densities. The plane on which the perspective plots are superimposed corresponds to the plane delimited by the first two axes (either PCo or NMDS). For brevity, we have not built bivariate density or perspective plots for the two character partitions.

Statistical protocols. Clustering and overdispersion of taxa in the bivariate scatterplots from PCoA and NMDS were estimated for fish and tetrapods combined, as well as for each of these two groups considered separately. The Ripley-Rasson method was again employed to estimate spatial bounds, but this time defining convex polygonal windows. For all species, as well as for fish and tetrapods independently, we calculated Ripley’s K function (Ripley 1979), describing the average probability of encountering neighbouring points as the distance from a focal point increases (e.g. Baddeley and Turner 2005; Hammer and Harper 2006). The K function for any given (observed) point distribution was first compared with the theoretical distribution that would be expected for a random (Poisson) scatter within the defined bounds of the observation window. If portions of the curve from the real data points lie above or below the theoretical curve (it can cross it repeatedly, depending upon the scales of spatial randomness), then the real data points are more or less clustered, respectively, at those distances than we would expect from a Poisson model. In order to assess the significance of these differences, we used Monte Carlo simulations of 999 random point distributions defining empirical confidence ‘envelopes’ around the theoretical curves. The portions of the observed curve occurring above or below the highest and lowest boundaries of these envelopes indicate that the

species are significantly more or less clustered at the corresponding scales, respectively. The K function and confidence envelopes are illustrated for both PCoA (combinations of first three axes; Fig. 7) and NMDS (first two dimensions; Fig. 8).

Character partition analyses

PAUP* v. 4.0b10 (Swofford 2002) was employed in parsimony analyses based upon two character set partitions of the Ruta (2011) data matrix. In all cases, the parsimony settings were as follows: heuristic searches; 5000 random stepwise addition sequences; tree bisection-reconnection branch-swapping algorithm (TBR), keeping a single tree in memory at the end of each addition sequence; TBR branch-swapping subsequently applied to all trees in memory, with the option of saving multiple trees (see Quicke *et al.* 2001); “amb-” option in effect (internal branches are collapsed if their minimum possible length is zero). The two character partitions were as follows: characters 1–89, 139–143, 151, and 153–156 for the anterior appendicular set; characters 90–138, 144–148, and 157 for the posterior appendicular set. Characters 149, 150, and 152 were general to both anterior and posterior sets, and were therefore excluded. Taxa that contained problematic amounts of missing data in one or both partitions were removed, leaving 27 terminals in the reduced dataset. The taxa removed included 11 fin-bearing (*Glyptolepis*, *Gogonasus*, *Marsdenichthys*, *Medoevia*, *Onychodus*, *Rhizodopsis*, *Rhizodus*, *Sauripterus*, *Sterropterygion*, *Strepsodus*, *Tiktaalik*) and six limb-bearing (or presumed limb-bearing) species (*Caerorhachis*, the Catskill humerus, *Hynerpeton*, *Ichthyostega*, *Pederpes*, *Silvanerpeton*).

In our data partition experiments, we first applied the standard incongruence length difference (ILD) test (e.g. Farris *et al.* 1995a, b) to the modified matrix. The ILD compares the length of the most parsimonious trees (MPTs) inferred from the entire character set with the sum of the lengths of the MPT(s) inferred from two (or more) character partitions analysed independently. The difference between these two lengths is the incongruence length difference (ILD), and could be attributed to

character conflict between the partitions. The test determines whether the difference is consistent with a null expectation, specifically whether the observed ILD value is within the 95 percent confidence interval of ILD values for data sets randomly partitioned in the same proportions as the original data (in this case, 99 pectoral and 55 pelvic characters). A large number (500) of random partitions was used to build a distribution of random ILDs.

In our second set of data partition experiments, we applied a test analogous to the ILD, but using tree-to-tree distance metrics in the place of tree length differences as a measure of incongruence. Two metrics were used: the symmetrical difference distance, RF (e.g. Robinson and Foulds 1981) and the maximum agreement subtree distance, d1 (e.g. Finden and Gordon 1985). For two trees, RF is calculated as the number of clades (or taxon partitions) that are unique to both. Similarly, d1 is defined as the number of taxa that have to be removed to obtain a maximum agreement subtree, i.e. a tree consisting of a subset of taxa for which all trees agree upon relationships after removing the smallest possible number of taxa. Specifically, we determined whether the relationships among taxa generated from the pectoral and pelvic data analysed independently differed significantly from the relationships among taxa generated from a large number of random, similarly sized character partitions. As multiple MPTs were often obtained, we calculated the distances between the two sets of trees as the mean of distances between nearest equivalent trees in these sets (e.g. Cobbett *et al.* 2007).

RESULTS

Phylogenetic analyses with character partitions

We present the results from the phylogenetic analyses first, as they assist in the discussion of the disparity results. Figure 1A shows the original strict consensus from Ruta (2011). Figure 1B-G illustrate 50 percent majority-rule and Adams consensus trees from maximum parsimony analyses

of the 27 terminals using the whole character set (minus characters 149, 150, and 152; see above) (Fig. 1B, E), and from the anterior (Fig. 1C, F) and posterior (Fig. 1D, G) character partitions, respectively. The majority-rule consensus shows the proportion of maximum parsimonious trees in which a specific node appears. We further chose the Adams consensus to provide an immediate characterization of unstable taxa in the most parsimonious trees. In an Adams consensus, such taxa are shifted to the most inclusive node in which they appear in any of the shortest trees. Analyses of the taxon-reduced matrix result in significant taxon rearrangements relative to the original tree, especially when only pelvic data are considered. When all characters are used for the analysis of the 27 terminals (Fig. 1B, E), two most parsimonious trees (condensed from four with the “amb-” option) are found (403 steps; global consistency index, CI = 0.5215; global retention index, RI = 0.6577). *Crassigyrinus* and baphetids (*Eucritta*, *Baphetes*) emerge, respectively, as the most basal and second most basal taxon after *Acanthostega*. *Tulerpeton* and whatcheeriids (*Ossinodus*, *Whatcheeria*) are sister groups and, together, they form the sister group to anthracosaurs (*Pholiderpeton*, *Proterogyrinus*, *Eoherpeton*, *Archeria*). All other limbed taxa preserve the same relative positions as in the trees from the original analysis (Fig. 1A). When pectoral characters are employed with the 27 terminals (Fig. 1C, F), 38 most parsimonious trees (condensed from 142 with the “amb-” option) are found (236 steps; CI = 0.6102; RI = 0.716). *Crassigyrinus* emerges the second basalmost taxon after *Acanthostega*, *Greererpeton* and *Tulerpeton* occur in a polytomy near the base of all subsequent limbed taxa, and whatcheeriids fail to appear as a clade in the Adams consensus. In the rest of the tree, microsaurids emerge as a clade, but neotridians and temnospondyls form overlapping arrays of taxa. ~~Considerable~~ Considerably more disruption takes place when pelvic characters are used alone (Fig. 1D, G; 99281 most parsimonious trees condensed from 397062 with the “amb-” option; 236 steps; CI = 0.5101; RI = 0.6178). By far the most unexpected result is the placement of several finned taxa among basal limbed taxa. Most major limbed clades are not retrieved in the Adams consensus, but a notable exception is represented by neotridians.

The ILD test returns a significant result ($p = 0.002$), suggesting differences in the nature of the signals associated with the pectoral and pelvic character sets (Fig. 2A). However, we urge caution in the interpretation of these results because the ILD test is known to be very (perhaps excessively) stringent, with a high type I error rate (see Planet 2006; Dolphin *et al.* 2000).

Our tests based on tree distances are, considered together, slightly more equivocal than the ILD test. Thus, whilst the RF test returns a significant result ($p = 0.002$) (Fig. 2B), the d1 test is marginally non-significant ($p = 0.062$) (Fig. 2C). Overall, we think it reasonable to conclude that there is significant incongruence in the pectoral and pelvic signals, as shown by numerous attributes of the MPTs generated from these two character partitions, such as are described above. We do not use such incongruence as an argument against combining partitions. Thus, we acknowledge the possibility of hidden support (Gatesy *et al.* 1999) and the value of total evidence (Kluge 1989), but utilise the tests here to explore the nature of the signal in the two partitions.

Disparity analyses

General remarks. The number of PCo axes explaining at least 90 per cent of the total variance is 36 when all characters are used, and 37 and 34, respectively, when anterior and posterior appendicular data are used. Tests of the differences in the means, medians, ranges, and variances of PCo scores for fish and tetrapods return significant results for the three character sets, regardless of the number of axes retained (~~Table~~ [Table S1](#); *Dryad Digital Repository*: <http://dx.doi.org/10.5061/dryad.1hr49/3>).

A global test of congruence among the three PCo distance matrices (built from all characters, and from the anterior and posterior appendicular characters) rejects the null hypothesis of complete incongruence (Friedman's $X^2 = 2461.929$; Kendall's $W = 0.868$; $p = 0.001$). The *post-hoc* tests show that each of the three matrices, in turn, is congruent with the other two (i.e. permutation tests for the Mantel statistics are significant; $p = 0.001$). Finally, the rank-based Mantel correlations show that

the degree of resemblance is highest for the two matrices of PCo distances generated from the full and the anterior appendicular sets (0.977), lowest for the matrices generated from the anterior and posterior appendicular sets (0.699), and intermediate for the matrices generated from the full and the posterior appendicular sets (0.722).

In general agreement with the results from CADM, the Procrustean superimposition method shows that the highest correlation between the configurations of taxa in morphospace occurs between the distance matrices from the full and anterior appendicular sets (Procrustes sum of squares = 0.01475; correlation in symmetric Procrustes rotation = 0.9926; $p = 0.001$). Intermediate value is found for the distance matrices from the full and posterior sets (Procrustes sum of squares = 0.0789; correlation in symmetric Procrustes rotation = 0.9597; $p = 0.001$). The lowest value is found for the distance matrices from the anterior and posterior sets (Procrustes sum of squares = 0.1016; correlation in symmetric Procrustes rotation = 0.9478; $p = 0.001$).

Rarefaction. In the vast majority of disparity indices investigated here (Fig. 3A, C, D), the median disparity in tetrapods using the whole data set is either marginally greater than or similar to that of fish over most comparable bootstrapped taxon sample sizes. At their greatest sample size, tetrapods are invariably more disparate than fish (except for the sum of variances) (Fig. 3B). Following Foote (1992), significant differences can be inferred when the mean or median values for the smaller group(s) lie(s) outside the 90 percent confidence intervals for the largest group. In our study, three indices (i.e. sum of ranges and the two root-products; Fig. 3A, C, D) do not differ significantly, at least for the maximum sample size of fish. The fourth index (sum of variances; Fig. 3B) shows that fish deviate more strongly from their own centroid (that is, they are more spread out relative to their own mean than tetrapods are). As would be expected for a variance-based index, the median values of the sum of variances change remarkably little with sample size for both groups. Based upon the two sum indices, we conclude that fish and tetrapods occupy morphospace regions of approximately

similar extent (such as is defined by the most eccentric data points), but within those regions their respective distributions differ.

With the anterior appendicular set, the rarefaction analyses are very similar to those derived from the entire data set in the case of the two sums (Fig. 4A, B). However, the two root-products (Fig. 4C, D) show tetrapods to be more disparate than fish (i.e. the median value for the maximum sample size of fish is outside the confidence interval for tetrapods).

Finally, the posterior appendicular set supports a higher median disparity in tetrapods than in fish, significantly so at all sample sizes greater than three for the sum of ranges (Fig. 4E) and 10 for the sum of variances (Fig. 4F). The occurrence of zero-value ranges and variances for scores on some PCo axes made it impossible to calculate product-based indices.

Morphospace plots. All two-dimensional plots using the first three PCo axes (for brevity, we show only those based upon the entire data set; Fig. 5A–C) are broadly similar, although the plot based upon the posterior set differs somewhat from those based upon the whole and the anterior sets. A two-dimensional NMDS is shown in Figure 5D. Various general features of both the PCo and the NMDS plots are discussed here, as they provide a semi-quantitative characterization of the data point distribution. In both plots, all fin-bearing taxa form a widespread and slightly discontinuous scatter. In addition, they are aligned for the most part along the second and third PCo axes (but there is little spread on the first PCo axis) in the first two plots (Fig. 5A, B), and do not overlap limb-bearing taxa. In the third plot (Fig. 5C), however, we note that fin-bearing taxa are almost equally spread on the second and third PCo axes; also, they surround the majority of limb-bearing taxa (with which they intersperse), and show a discontinuous distribution. Most limb-bearing taxa display a compact distribution in the three PCo plots, with little spread along the first three PCo axes. Five taxa (i.e. *Acanthostega*, *Caerorhachis*, Catskill humerus, *Hynerpeton*, *Ichthyostega*) are consistently well separated from other limb-bearing taxa. A sixth taxon, *Ossinodus*, emerges as an outlier in the first and third plots (Fig. 5A, C). In the compact cluster, we find mostly Carboniferous

and Permian taxa (*Tulerpeton* is an exception). In agreement with conclusions from comparative anatomical and phylogenetic analyses, *Acanthostega*, *Ichthyostega*, *Hynerpeton* and the Catskill humerus plot out in morphospace between elpistostegids and post-Devonian limbed taxa. Finally, the pattern of the NMDS plot (Fig. 5D) is very similar to that from the first two PCo plots (Fig. 5A, B): finned and limbed taxa are separate, and the latter are more densely clustered than the former.

Spatial analyses

Density plots. Both density (Fig. 6A, B) and perspective (Fig. 6C, D) plots show the discontinuity between fish and tetrapods. In the interest of brevity, we do not show plots associated with taxon distributions built from the data partitions. However, these plots also indicate heterogeneous taxon distributions in morphospace and the clear separation between fish and tetrapods (see above). One outstanding difference between the density and perspective plots related to PCoA (Fig. 6A, C) and those related to NMDS (Fig. 6B, D) is that fish are more uniformly distributed in the NMDS-related plots, whereas the PCoA-related plots show marginally greater discontinuity and two very slightly distinct peaks for fish.

Data point patterns. Ripley's K function was used to investigate the spatial distributions of taxa in morphospace for different inter-taxon distance scales. The K function expresses the average point density (or probability) as a function of the distance from every point (Hammer and Harper 2006) (Fig. 7: 'iso' line). By comparing the observed point distribution with a theoretical distribution from a random Poisson model (Fig. 7: 'theo' line), it is possible to determine whether taxa are under- or over-dispersed at different distance scales. The addition of a confidence envelope using 999 Monte Carlo-simulated random processes (Fig. 7; area between 'high' and 'low' lines) shows whether or not these differences are significant. When all fish and tetrapods are considered simultaneously, the observed curves occur significantly above the theoretical ones for all bivariate combinations of the

first three PCo axis coordinates (Fig. 7A, D, G). This indicates that the taxa are more clustered than expected, regardless of scale. Analysis of the tetrapod distribution alone yield similar, though less marked, results (Fig. 7C, F, I). By contrast, fish distribution is similar to a Poisson distribution on the plane delimited by PCo axes 1 and 2 (Fig. 7B), and statistically indistinguishable from a Poisson (though greater than it) on the planes delimited by PCo axes 1 and 3, and 2 and 3 (Fig. 7E, H).

Very similar results to those based upon PCoA were obtained for the two-dimensional NMDS plot (Fig. 8). Except at very small distances, fish and tetrapod taxa together are significantly more clustered than expected (Fig. 8A). Fish are more dispersed than expected, though not significantly so (Fig. 8B). Conversely, tetrapods are significantly more clustered than expected at mid to large distances, but not at small to mid distances (Fig. 8C).

Comments on a few key taxa

Each of the taxa bracketing closely the fish-tetrapod internode (*Panderichthys*; *Tiktaalik*; Catskill humerus; *Acanthostega*; *Ichthyostega*) was tested for phenetic proximity to all remaining fish or all remaining tetrapods (Mann-Whitney U tests). *Acanthostega* ($U = 26$; $p < 0.001$), *Ichthyostega* ($U = 1$; $p < 0.001$) and the Catskill humerus ($U = 107$; $p = 0.009$) were significantly closer to tetrapods than they were to fish. In the case of *Panderichthys* ($U = 162.5$; $p = 0.200$) and *Tiktaalik* ($U = 184$; $p = 0.632$), the tests failed to reject the null hypothesis of no differences in the distributions of their distances from fish and from tetrapods. This is expected, given the mosaic of transitional traits of both genera, witnessed by their phylogenetic position (Daeschler *et al.* 2006; Shubin *et al.* 2006; Boisvert 2009; Ruta 2011).

The Catskill humerus raises some interesting questions. As no other part of the anatomy of this animal is known, it is difficult to place phylogenetically. In its construction, it is certainly more derived than the humeri of fin-bearing taxa (including those of *Panderichthys* and *Tiktaalik*). On the other hand, it is more primitive than the humeri of *Acanthostega* and *Ichthyostega* (see Coates *et al.*

2004). Parsimony with equally weighted characters is inconclusive (Text-fig. 1A; see Ruta 2011).

In order to quantify overall phenetic proximity of the Catskill humerus to fish and tetrapods, we repeated the distance tests using solely the anterior appendicular data and found, once again, that the Catskill humerus is phenetically closer to the limb-bearing taxa than to the fin-bearing taxa ($U = 112$; $p = 0.0129$).

A note on the distribution of unique character-states

We report results of two tests devised to assess whether the unique states in the data matrix occur randomly. We recall that these tests use the distribution of those states and evaluate its departure (significant or otherwise) from a Poisson distribution. In each of the two graphs in Figure 9, grey bars are histograms representing the square root of number of occurrences of uniquely expressed states (i.e. instances of assignment of state 0, state 1, state 2, ..., state n to a single taxon). The grey dots are expected counts based on a Poisson distribution. The histograms are displaced vertically in order to match the expected counts. This displacement can be thought of as the amount of departure from the Poisson distribution. Both maximum likelihood and chi-square were used as optimization criteria, treating each autapomorphy count as a category, and inferring the expected counts from the theoretical curves. A goodness-of-fit test for the Poisson distribution returns significant results, both with the maximum likelihood (likelihood ratio = 15.591; $p = 0.004$) and with the chi-square optimization criteria (Pearson = 23.216; $p < 0.001$). These results suggest that the distribution of the unique states in the data matrix is non-random.

DISCUSSION

Phylogeny

Two of the partition tests show that there are significant differences in the nature of the signals from pectoral and pelvic characters. The highly significant result for the ILD ($p = 0.002$) is unsurprising given that the test is notoriously difficult to pass. Furthermore, the ILD can return a significant result because of differing signal strengths in the partitions, as well as because the trees imply different sets of relationships (e.g. Planet 2006; Dolphin *et al.* 2000). In our study, however, an identical p -value is obtained from the tree distance-based analogue of the ILD test using RF, and a similar test using d1 was only marginally non-significant ($p = 0.062$). We conclude that inferred relationships differ significantly in the two partitions (see Adams consensus trees: Fig. 1E, F, G). Moreover, the trees built from anterior appendicular data are more similar to those from the entire, taxon-reduced dataset than are the trees built from posterior appendicular data (although the number of characters in the pectoral partition is greater). The differences in p -values arise because the tree-to-tree distance metrics, RF and d1, measure different properties of the trees. Thus, we recall that the RF distance counts the number of nodes unique to both trees; it follows that such a distance is particularly sensitive to translocations of distant branches, even when the size of those branches (number of subtended terminals) is small. The d1 distance calculates the number of taxa that would need to be pruned from two trees in order to make them identical. As such, it is almost impervious to the distance of the translocations, but is more sensitive to the number of taxa involved in branch swaps (e.g. Cobbett *et al.* 2007). Our results suggest that the pectoral and pelvic characters imply trees that differ in both respects.

Disparity

Despite the diverse appendicular constructions of limb-bearing taxa, our sample is overall no more disparate than that of fin-bearing taxa when the full character set is considered. These results may appear to be counterintuitive at first, given the sheer diversity of tetrapod limbs and girdles for both Devonian stem-group and Carboniferous stem- and basal crown-group tetrapods (e.g. Clack 2012).

However, fins and girdles in representatives of the tetrapod stem-group (rhizodonts, osteolepiforms, and elpistostegids) also reveal a remarkable structural (and presumably functional) variety (e.g. Coates *et al.* 2002, 2007, 2008; Long *et al.* 2006; Friedman *et al.* 2007; Shubin *et al.* 2006; Holland and Long 2009; Zhu and Yu 2009; Friedman and Brazeau 2010).

~~A Possible explanation~~ for the observed disparity patterns are detailed briefly. First, such patterns may be due to the is that fact that several character-states occur ubiquitously – in some combination or another – in ~~many~~ several basal crown tetrapod clades. Ruta (2011) provided a preliminary and indirect characterization of this phenomenon, by showing that the rates of character-state change per branch, (particularly internal branches along the ‘backbone’ ~~(-or-~~ stem lineage) of the tree, decrease remarkably in post-Devonian groups. An alternative explanation (as yet untested, pending completion of a new character matrix) invokes character exhaustion (see Wagner 2000; Wagner *et al.* 2006), i.e. a decreasing frequency of novel character-states in descendants relative to ancestors. Exhaustion would result in character-state changes for tetrapods being distributed over a smaller number of characters than for fish. A third explanation might be that the conquest of land imposed functional constraints on the range of limb variability, perhaps as a result of the addition of a novel functional role – body weight support – to propulsion (Professor Mike Benton, personal communication to the authors, 2015; see also Clack 2012). An analogy can be drawn with the functional constraints imposed by flight on the skeleton of bird fore-limbs (see also Puttick *et al.* 2014). Finally, it might be hypothesized that limb characters, especially among basal crown tetrapod clades, have been scrutinised to a much lesser degree than those in taxa spanning the fin-limb transition proper. Thus, uneven efforts in delimitation of morphological conditions across all limbed clades and grades might have resulted in a character coding exhaustion artefact (Professor Mike Benton, personal communication to the authors, 2015).

For three of the disparity indices (the two root-products, and the sum of variances), there are no *significant* differences between fish and tetrapods, although tetrapods are slightly more disparate at most sample sizes. Conversely, the fourth index – sum of variances – is significantly greater for

Formatted: Font: Italic

fish. The additive (sum) indices take no account of the distribution of univariate ranges or variances (e.g. $5 + 5 + 5 = 7 + 7 + 1$), whereas the multiplicative (product) indices do (e.g. $5 \times 5 \times 5 > 7 \times 7 \times 1$). Thus, group A can have a higher sum *but* a lower product of univariate measures than group B (e.g. A: $12 + 2 + 2 = 16$ and $12 \times 2 \times 2 = 48$; B: $4 + 4 + 4 = 12$ and $4 \times 4 \times 4 = 64$). Higher products can thus be achieved by distributions with more similar ‘spread’ on all of the axes or coordinates. In our study, the bounds of morphospace envelope (range) are similar for fish and tetrapods whether such bounds are calculated by multiplication or addition. In contrast, the results for variances (i.e. mean squared difference from the group mean) differ for products (fish and tetrapods are similar) and sums (fish are greater than tetrapods). This implies that the differences among fish taxa are relatively unevenly distributed across coordinates (the difference between products and sums). Tetrapods appear to hold the bounds of the morphospace envelope open with a smaller number of ‘outliers’, to exhibit greater local clustering, and to concentrate the differences between them on a smaller number of axes. This is consistent with the results from the *K* function analyses, whereby tetrapods are significantly clustered at all spatial scales using the first three PCo axes, but the distribution of fish is statistically indistinguishable from random.

The greater dispersion of fish from their mean morphology *could* be interpreted as an artefact of poor taxon sampling. In short, we may have sampled fish taxa more selectively than tetrapod taxa because we included only those fish species for which detailed appendicular data are available. This could imply that the data matrix samples variation *between* different plesions on the piscine stretch of the tetrapod stem efficiently, whilst simultaneously undersampling variation *within* each plesion. Unfortunately, this possibility cannot be tested at present, as the relevant data are not available. If poorly-preserved (omitted) fish taxa were distributed at random along the tetrapod stem, then we would expect to sample most of the major plesions. In addition, the variance structure of our sample should reflect that of the original distribution. For example, if we had morphospace clusters (within plesions) with a modest scatter of taxa in between, then this would be seen in the distribution of fish in morphospace. However, if omitted fish were not distributed randomly (e.g. due to a preservation

or collection bias for some plesions), then a much more rarefied sample (large, empty morphospace areas) would likely result from the omission of whole plesions (or large proportions of those). The pattern we retrieve is therefore likely to be partly real.

Acknowledgements. We thank Drs Zerina Johanson (Manuscript Editor; The Natural History Museum) and Sally Thomas (Publications Officer; The Palaeontological Association) for their invaluable assistance with the production of this paper, and Professor Mike Benton (University of Bristol) and an anonymous referees XXX and XXX for their useful discussions, comments, and critical input, and supportive remarks. This ~~research work~~ is supported by the Natural Environment Research Council (Advanced Research Fellowship NE/F014872/1 to MR) and the Leverhulme Trust (Grant F/00351/Z to MAW).

DATA ARCHIVING STATEMENT

Data for this study are available in the *Dryad Digital Repository*;
<http://dx.doi.org/10.5061/dryad.1hr49>

REFERENCES

- ADAMOWICZ, S. J., PURVIS, A. and WILLS, M. A. 2008. Increasing morphological complexity as a major evolutionary trend in the Crustacea. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 4786–4791.
- AHLBERG, P. E. 2011. Humeral homology and the origin of the tetrapod elbow: a reinterpretation of the enigmatic specimens ANSP 21350 and GSM 104536. 17–29. *In* BARRETT, P. M. and MILNER, A. R. (eds). *Studies on fossil tetrapods. Special Papers in Palaeontology*, **86**. Wiley-Blackwell, Oxford, 280 pp.

Formatted: Font: Bold

Formatted: Font: Times New Roman, Bold

Formatted: Line spacing: Double

Formatted: Font: Times New Roman

Formatted: Line spacing: Double

Formatted: Font: Italic

Formatted: Font: Times New Roman

Formatted: Font: Times New Roman, 12 pt

Formatted: English (United Kingdom)

- ANDERSON, J. S. 2008. Focal review: the origin(s) of modern amphibians. *Evolutionary Biology*, **35**, 231–247.
- , SMITHSON, T., MANSKY, C. F., MEYER, T. and CLACK, J. A. 2015. A -diverse tetrapod fauna at the base of ‘Romer’s Gap’. *PLoS One*, **10** (e0125446), 1–27.
- ANDERSON, M. J. 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, **26**, 32–46.
- AUSICH, W. I. and DELINE, B. 2012. Macroevoolutionary transition in crinoids following the Late Ordovician extinction event (Ordovician to Early Silurian). *Palaeogeography, Palaeoclimatology, Palaeoecology*, **361**, 38–48.
- BADDELEY, A. and TURNER, R. 2005. Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software*, **12**, 1–42.
- BAPST, D. W., BULLOCK, P. C., MELCHIN, M. J., SHEETS, H. D. and MITCHELL, C. E. 2012. Graptoloid diversity and disparity became decoupled during the Ordovician mass extinction. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 3428–3433.
- BENSON, R. B. J., EVANS, M. and DRUCKENMILLER, P. S. 2012. High diversity, low disparity and small body size in plesiosaurs (Reptilia, Sauropterygia) from the Triassic–Jurassic boundary. *PLoS One*, **7** (e31838), 1–15.
- BENTON, M. J. 2015. Exploring macroevolution using modern and fossil data. *Proceedings of the Royal Society of London, Biological Sciences*, **282** (e20150569), 1–10.
- BOISVERT, C. A. 2009. The humerus of *Panderichthys* in three dimensions and its significance in the context of the fish-tetrapod transition. *Acta Zoologica*, **90** (Issue Supplement S1), 297–305.
- , MARK-KURIK, E. and AHLBERG, P. E. 2008. The pectoral fin of *Panderichthys* and the origin of digits. *Nature*, **456**, 636–638.

- BRIGGS, D. E. G., FORTEY, R. A. and WILLS, M. A. 1992. Morphological disparity in the Cambrian. *Science*, **256**, 1670–1673.
- BRUSATTE, S. L., BENTON, M. J., RUTA, M. and LLOYD, G. T. 2008a. Superiority, competition, and opportunism in the evolutionary radiation of dinosaurs. *Science*, **321**, 1485–1488.
- , —, —, — 2008b. The first 50 million years of dinosaur evolution: macroevolutionary pattern and morphological disparity. *Biology Letters*, **4**, 733–736.
- , BUTLER, R. J., PRIETO-MÁRQUEZ, A. and NORELL, M. A. 2012. Dinosaur morphological diversity and the end-Cretaceous extinction. *Nature Communications*, **3**, 1–8.
- , MONTANARI, S., YI, H.-Y. and NORELL, M. A. 2011. Phylogenetic corrections for morphological disparity analysis: new methodology and case studies. *Paleobiology*, **37**, 1–22.
- BUTLER, R. J., BRUSATTE, S. L., ANDRES, B. and BENSON, R. B. 2012. How do geological sampling biases affect studies of morphological evolution in deep time? A case study of pterosaur (Reptilia: Archosauria) disparity. *Evolution*, **66**, 147–162.
- CÁCERES, M. DE, FONT, X., OLIVA, F. and VIVES, S. 2007. Ginkgo, a program for non-standard multivariate fuzzy analysis. *Advances in Fuzzy Sets and Systems*, **2**, 41–56.
- CAILLIEZ, F. 1983. The analytical solution of the additive constant problem. *Psychometrika*, **48**, 305–308.
- CALLIER, V., CLACK, J. A. and AHLBERG, P. E. 2009. Contrasting developmental trajectories in the earliest known tetrapod forelimbs. *Science*, **324**, 364–367.
- CIAMPAGLIO, C. N., KEMP, M. and MCSHEA, D. W. 2001. Detecting changes in morphospace occupation patterns in the fossil record: characterization and analysis of measures of disparity. *Paleobiology*, **27**, 695–715.
- CISNEROS, J. C. and RUTA, M. 2010. Morphological diversity and biogeography of procolophonids (Amniota: Parareptilia). *Journal of Systematic Palaeontology*, **8**, 607–625.
- CLACK, J. A. 2009. The fin to limb transition: new data, interpretations, and hypotheses from

- paleontology and developmental biology. *Annual Review of Earth and Planetary Sciences*, **37**, 163–179.
- 2012. *Gaining ground: the origin and evolution of tetrapods*. Indiana University Press, Bloomington, 544 pp.
- CLARKE, K. R. 1993. Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology*, **18**, 117–143.
- COATES, M. I., JEFFERY, J. E. and RUTA, M. 2002. Fins to limbs: what the fossils say. *Evolution and Development*, **4**, 390–401.
- , RUTA, M. and FRIEDMAN, M. 2008. Ever since Owen: changing perspectives on the early evolution of tetrapods. *Annual Review of Ecology, Evolution, and Systematics*, **39**, 571–592.
- , — and WAGNER, P. J. 2007. Using patterns of fin and limb phylogeny to test developmental-evolutionary scenarios. 245–261. *In* Novartis Foundation Symposium, No. 284. *Tinkering: the microevolution of development*. John Wiley & Sons, Chichester, 304 pp.
- , SHUBIN, N. H. and DAESCHLER, E. B. 2004. Response to Comment on “The Early Evolution of the Tetrapod Humerus”. *Science*, **305**, 1715.
- COBBETT, A., WILKINSON, M. and WILLS, M. A. 2007. Fossils impact as hard as living taxa in parsimony analyses of morphology. *Systematic Biology*, **56**, 753–766.
- DAESCHLER, E. B., SHUBIN, N. H. and JENKINS, F. A. Jr. 2006. A Devonian tetrapod-like fish and the evolution of the tetrapod body plan. *Nature*, **440**, 757–763.
- DOLPHIN, K., BELSHAW, R., ORME, C. D. L. and QUICKE, D. L. J. 2000. Noise and incongruence: interpreting results of the incongruence length difference test. *Molecular Phylogenetics and Evolution*, **17**, 401–406.
- ERWIN, D. H. 2015. A public goods approach to major evolutionary radiations. *Geobiology*, **13**, 308–315.
- FARRIS, J. S., KÄLLERSJÖ, M., KLUGE, A. G. and BULT, C. 1995a. Constructing a significance test for incongruence. *Systematic Biology*, **44**, 570–572.

- , —, —, — 1995*b*. Testing significance of incongruence. *Cladistics*, **10**, 315–319.
- FINDEN, C. R. and GORDON, A. D. 1985. Obtaining common pruned trees. *Journal of Classification*, **2**, 255–276.
- FOTH, C., BRUSATTE, S. L. and BUTLER, R. J. 2012. Do different disparity proxies converge on a common signal? Insights from the cranial morphometrics and evolutionary history of Pterosauria (Diapsida: Archosauria). *Journal of Evolutionary Biology*, **25**, 904–915.
- FRIEDMAN, M. and BRAZEAU, M. D. 2010. A reappraisal of the origin and basal radiation of the Osteichthyes. *Journal of Vertebrate Paleontology*, **30**, 36–56.
- , COATES, M. I. and ANDERSON, P. J. 2007. First discovery of a primitive coelacanth fin fills a major gap in the evolution of lobed fins and limbs. *Evolution and Development*, **9**, 329–337.
- GATESY, J., O'GRADY, P. and BAKER, R. H. 1999. Corroboration among data sets in simultaneous analysis: hidden support for phylogenetic relationships among higher level artiodactyl taxa. *Cladistics*, **15**, 271–313.
- GERBER, S. 2013. On the relationship between the macroevolutionary trajectories of morphological integration and morphological disparity. *PLoS One*, **8** (e63913), 1–8.
- GROSSNICKLE, D. M. and POLLY, P. D. 2013. Mammal disparity decreases during the Cretaceous angiosperm radiation. *Proceedings of the Royal Society of London, Biological Sciences*, **280** (e20132110), 1–8.
- HALL, B. K. (ed.). 2007. *Fins into limbs: evolution, development, and transformation*. The University of Chicago Press, Chicago, 344 pp.
- HAMMER, Ø. and HARPER, D. 2006. *Paleontological data analysis*. Blackwell Publishing, Oxford, 351 pp.
- , — and RYAN, P. D. 2001. PAST: paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*, **4**, 1–9.

- HETHERINGTON, A. J., SHERRATT, E., RUTA, M., WILKINSON, M., DELINE, B. and DONOGHUE, P. C. J. 2015. Do cladistic and morphometric data capture common patterns of morphological disparity? *Palaeontology*, **58**, 393–399.
- HOLLAND, T. and LONG, J. A. 2009. On the phylogenetic position of *Gogonasus andrewsae* Long 1985, within the Tetrapodomorpha. *Acta Zoologica*, **90**, 285–296.
- HOLMES, R. B. 2000. Palaeozoic temnospondyls. 1081–1120. In HEATWOLE, H. and CARROLL, R. L. (eds). *Amphibian biology, volume 4. Palaeontology, the evolutionary history of amphibians*. Surrey Beatty and Sons Ltd., Chipping Norton, Australia, 1496 pp.
- HUGHES, M., GERBER, S. and WILLS, M. A. 2013. Clades reach highest morphological disparity early in their evolution. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 13875–13879.
- JOHANSON, Z., JOSS, J., BOISVERT, C. A., ERICSSON, R., SUTIJA, M. and AHLBERG, P. E. 2007. Fish fingers: digit homologues in sarcopterygian fish fins. *Journal of Experimental Zoology Part B, Molecular and Developmental Evolution*, **308**, 757–768.
- KLUGE, A. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). *Systematic Biology*, **38**, 7–25.
- LARGET, B. and SIMON, D. L. 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Molecular Biology and Evolution*, **16**, 750–759.
- LEGENDRE, P. and LAPOINTE, F.-J. 2004. Assessing the congruence among distance matrices: single malt Scotch whiskies revisited. *Australian and New Zealand Journal of Statistics*, **46**, 615–629.
- and LEGENDRE, L. *Numerical ecology*. Elsevier Science, Amsterdam, 853 pp.
- LOFGREN, A. S., PLOTNICK, R. E. and WAGNER, P. J. 2003. Morphological diversity of Carboniferous arthropods and insights on disparity patterns through the Phanerozoic. *Paleobiology*, **29**, 349–368.
- LONG, J. A., YOUNG, G. C., HOLLAND, T., SENDEN, T. J. and FITZGERALD, E. M. G. 2006.

- An exceptional Devonian fish from Australia sheds light on tetrapod origins. *Nature*, **444**, 199–202.
- LUCY, D., AYKROYD, R. G. and POLLARD, A. M. 2002. Nonparametric estimation of age calibration. *Applied Statistics*, **51**, 183–196.
- MANTEL, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.
- MELORO, C. and JONES, M. E. H. 2012. Tooth and cranial disparity in the fossil relatives of *Sphenodon* (Rhynchocephalia) dispute the persistent “living fossil” label. *Journal of Evolutionary Biology*, **25**, 2194–2209.
- MEYER, D., Zeileis, A. and Hornik, K. 2015. vcd: Visualizing Categorical Data. R package version 1.4-0.
- NEENAN, J. M., RUTA, M., CLACK, J. A. and RAYFIELD, E. J. 2014. Feeding biomechanics in *Acanthostega* and across the fish-tetrapod transition. *Proceedings of the Royal Society of London: Biological Sciences*, **281** (e20132689), 1–8.
- OKSANEN, J., BLANCHET, F. G., KINDT, R., LEGENDRE, P., MINCHIN, P. R., O'HARA, R. B., SIMPSON, G. L., SOLYMOS, P., HENRY, M., STEVENS, H. and WAGNER, H. 2013. vegan: Community Ecology Package. R package version 2.0-10.
- PARADIS, E., CLAUDE, J. and STRIMMER, K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.
- PERES-NETO, P. and JACKSON, D. 2001. How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia*, **129**, 169–178.
- PIERCE, S. E., CLACK, J. A. and HUTCHINSON, J. R. 2012. Three-dimensional limb joint mobility in the early tetrapod *Ichthyostega*. *Nature*, **486**, 523–526.
- ~~PIERCE, S. E.~~, AHLBERG, P. E., HUTCHINSON, J. R., MOLNAR, J. L., SANCHEZ, S., TAFFOREAU, P. and CLACK, J. A. 2013. Vertebral architecture in the earliest stem

tetrapods. *Nature*, **494**, 226–230.

—, CLACK, J. A. and HUTCHINSON, J. R. 2012. Three-dimensional limb joint mobility in the early tetrapod *Ichthyostega*. *Nature*, **486**, 523–526.

PLANET, P. J. 2006. Tree disagreement: measuring and testing incongruence in phylogenies.

Journal of Biomedical Informatics, **39**, 86–102.

PRENTICE, K. C., RUTA, M. and BENTON, M. J. 2011. Evolution of morphological disparity in pterosaurs. *Journal of Systematic Palaeontology*, **9**, 337–353.

QUICKE, D. L. J., TAYLOR, J. and PURVIS, A. 2001. Changing the landscape: a new strategy for estimating large phylogenies. *Systematic Biology*, **50**, 60–66.

RIPLEY, B. D. 1979. Tests of ‘randomness’ for spatial point patterns. *Journal of the Royal Statistical Society, Series B*, **41**, 368–374.

— and RASSON, J.-P. 1977. Finding the edge of a Poisson forest. *Journal of Applied Probability*, **14**, 483–491.

ROBINSON, D. and FOULDS, L. 1981. Comparison of phylogenetic trees. *Mathematical Biosciences*, **53**, 131–147.

RUTA, M. 2009. Patterns of morphological evolution in major groups of Palaeozoic Temnospondyli (Amphibia: Tetrapoda). 91–120. In RUTA, M., CLACK, J. A. and MILNER, A. C. (eds). *Patterns and processes in early vertebrate evolution. Special Papers in Palaeontology*, **81**. Wiley-Blackwell, Oxford, 173 pp.

— 2011. Phylogenetic signal and character compatibility in the appendicular skeleton of early tetrapods. 31–43. In BARRETT, P. M. and MILNER, A. R. (eds). *Studies on fossil tetrapods. Special Papers in Palaeontology*, **86**. Wiley-Blackwell, Oxford, 280 pp.

— and BOLT, J. R. 2008. The brachyopoid *Hadrokkosaurus bradyi* from the early Middle Triassic of Arizona, and a phylogenetic analysis of lower jaw characters in temnospondyl amphibians. *Acta Palaeontologica Polonica*, **53**, 579–592.

- and CLACK, J. A. 2006. A review of *Silvanerpeton miripedes*, a stem amniote from the Lower Carboniferous of East Kirkton, West Lothian, Scotland. *Transactions of the Royal Society of Edinburgh: Earth Sciences*, **97**, 31–63.
- and COATES, M. I. 2007. Dates, nodes and character conflict: addressing the lissamphibian origin problem. *Journal of Systematic Palaeontology*, **5**, 69–122.
- , COATES, M. I. and QUICKE, D. L. J. 2003. Early tetrapod relationships revisited. *Biological Reviews of the Cambridge Philosophical Society*, **78**, 251–345.
- , WAGNER, P. J. and COATES, M. I. 2006. Evolutionary patterns in early tetrapods. I. Rapid initial diversification followed by decrease in rates of character change. *Proceedings of the Royal Society of London: Biological Sciences*, **273**, 2107–2111.
- , ANGIELCZYK, K. D., FRÖBISCH, J. AND BENTON, M. J. 2013a. Decoupling of morphological disparity and taxic diversity during the adaptive radiation of anamniotes. *Proceedings of the Royal Society of London: Biological Sciences*, **280** (e20131071), 1–9.
- and BOLT, J. R. 2008. The brachyopoid *Hadrokosaurus bradyi* from the early Middle Triassic of Arizona, and a phylogenetic analysis of lower jaw characters in temnospondyl amphibians. *Acta Palaeontologica Polonica*, **53**, 579–592.
- , BOTHA-BRINK, J., MITCHELL, S. A. and BENTON, M. J. 2013b. The radiation of cynodonts and the ground plan of mammalian morphological diversity. *Proceedings of the Royal Society of London: Biological Sciences*, **280** (e20131865), 1–10.
- and CLACK, J. A. 2006. A review of *Silvanerpeton miripedes*, a stem amniote from the Lower Carboniferous of East Kirkton, West Lothian, Scotland. *Transactions of the Royal Society of Edinburgh: Earth Sciences*, **97**, 31–63.
- and COATES, M. I. 2007. Dates, nodes and character conflict: addressing the lissamphibian origin problem. *Journal of Systematic Palaeontology*, **5**, 69–122.

- , — and QUICKE, D. L. J. 2003. Early tetrapod relationships revisited. *Biological Reviews of the Cambridge Philosophical Society*, **78**, 251–345.
- , PISANI, D., LLOYD, G. T. and BENTON, M. J. 2007. A supertree of Temnospondyli: cladogenetic patterns in the most species-rich group of early tetrapods. *Proceedings of the Royal Society of London: Biological Sciences*, **274**, 3087–3095.
- , WAGNER, P. J. and COATES, M. I. 2006. Evolutionary patterns in early tetrapods. I. Rapid initial diversification followed by decrease in rates of character change. *Proceedings of the Royal Society of London: Biological Sciences*, **273**, 2107–2111.
- SHUBIN, N. H. 2009. *Your inner fish*. Penguin, London, 256 pp.
- SHUBIN, N. H., DAESCHLER, E. B. and COATES, M. I. 2004. The early evolution of the tetrapod humerus. *Nature*, **304**, 90–93.
- , — and JENKINS, F. A. Jr. 2006. The pectoral fin of *Tiktaalik roseae* and the origin of the tetrapod limb. *Nature*, **440**, 764–771.
- , TABIN, C. and CARROLL, S. 2009. Deep homology and the origins of evolutionary novelty. *Nature*, **457**, 818–823.
- SMITH, A. J., ROSARIO, M. V., EITING, T. P. and DUMONT, E. R. 2014. Joined at the hip: linked characters and the problem of missing data in studies of disparity. *Evolution*, **68**, 2386–2400.
- SMITHSON, T. R. and CLACK, J. A. 2013. Tetrapod appendicular skeletal elements from the Early Carboniferous of Scotland. *Comptes Rendus Palevol*, **12**, 405–417.
- STUBBS, T. L., PIERCE, S. E., RAYFIELD, E. J. and ANDERSON, P. S. L. 2013. Morphological and biomechanical disparity of crocodile-line archosaurs following the end-Triassic extinction. *Proceedings of the Royal Society of London: Biological Sciences*, **280** (e20131940), 1–9.
- SWOFFORD, D. L. 2002. *PAUP*: phylogenetic analysis using parsimony (and other methods) 4.0 beta*. Sinauer Associates, Sunderland, Massachusetts.

- THORNE, P. M., RUTA, M. and BENTON, M. J. 2011. Resetting the evolution of marine reptiles at the Triassic-Jurassic boundary. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 8339–8344.
- TRAN, L. A. P. 2014. The role of ecological opportunity in shaping disparate diversification trajectories in a bicontinental primate radiation. *Proceedings of the Royal Society of London: Biological Sciences*, **281** (e20131979), 1–9.
- VAN VALEN, L. 1974. Multivariate structural statistics in natural history. *Journal of Theoretical Biology*, **45**, 235–247.
- VILLIER, L. and KORN, D. 2004. Morphological disparity of ammonoids and the mark of Permian mass extinctions. *Science*, **306**, 264–266.
- WAGNER, P. J. 2000. Exhaustion of morphologic character states among fossil taxa. *Evolution*, **54**, 365–386.
- , RUTA, M. and COATES, M. I. 2006. Evolutionary patterns in early tetrapods. II. Differing constraints on available character space among clades. *Proceedings of the Royal Society of London: Biological Sciences*, **273**, 2113–2118.
- WAND, M. P. and JONES, M. C. 1995. *Kernel smoothing*. Chapman and Hall, London, 224 pp.
- WILLS, M. A. 1998a. Crustacean disparity through the Phanerozoic: comparing morphological and stratigraphic data. *Biological Journal of the Linnean Society*, **65**, 455–500.
- 1998b. Cambrian and Recent disparity: the picture from priapulids. *Palaeobiology*, **24**, 177–199.
- 2001. Morphological disparity: a primer. 55–144. In ADRAIN, J. M., EDGECOMBE, G. D. and LIEBERMAN, B. S. (eds). *Fossils, phylogeny, and form. An analytical approach*. Kluwer Academic / Plenum Publishers, New York, 402 pp.
- , BRIGGS, D. E. G. and FORTEY, R. A. 1994. Disparity as an evolutionary index: a comparison of Cambrian and Recent arthropods. *Paleobiology*, **20**, 93–130.
- , GERBER, S., RUTA, M. and HUGHES, M. 2012. The disparity of priapulid, archaeopriapulid and palaeoscolecoid worms in the light of new data. *Journal of Evolutionary Biology*, **25**,

2056–2076.

- WOLTERING, J. M. and DUBOULE, D. 2010. The origin of digits: expression patterns versus regulatory mechanisms. *Developmental Cell*, **8**, 526–532.
- XUE, J., HUANG, P., RUTA, M., BENTON, M. J., HAO, S., XIONG, C., WANG, D., CASCALES-MIÑANA, B., WANG, Q. and LIU, L. 2015. Stepwise evolution of Paleozoic tracheophytes from South China: contrasting leaf disparity and taxic diversity. *Earth-Science Reviews*, **148**, 77–93.
- YOUNG, M. T., BRUSATTE, S. L., RUTA, M. and ANDRADE, M. B. DE 2010. The evolution of Metriorhynchoidea (Mesoeucrocodylia, Thalattosuchia): an integrated approach using geometric morphometrics, analysis of disparity, and biomechanics. *Zoological Journal of the Linnean Society*, **158**, 801–859.
- ZELDITCH, M. L., SWIDERSKI, D. L., SHEETS, H. D. and FINK, W. L. 2004. Geometric morphometric for biologists: a primer. Academic Press, New York, 416 pp.
- ZHANG, J., WAGH, P., GUAY, D., SANCHEZ-PULIDO, L., PADHI, B. K., KORZH, V., ANDRADE-NAVARRO, M. A. and AKIMENKO, M.-A. 2010. Loss of fish actinotrichia proteins and the fin-to-limb transition. *Nature*, **466**, 234–237.
- ZHU, M. and YU, X. 2009. Stem sarcopterygians have primitive polybasal fin articulation. *Biology Letters*, **5**, 372–375.

TEXT-FIG. 1. Early tetrapod relationships based on maximum parsimony analyses of appendicular data. A, strict consensus of 120 tree from the analysis of the full data matrix in Ruta (2011; 44 taxa, 157 characters); B, E, 50 percent majority-rule and Adams consensus of taxonomically reduced taxon set (27 out of 44 original taxa; see text for details) with 154 characters (removal of characters 149, 150, and 152 from original matrix; see text for details); C, F, 50 percent majority-rule and Adams consensus of taxonomically reduced taxon set using anterior appendicular characters (1–89, 139–143, 151, and 153–156); D, G, 50 percent majority-rule and Adams consensus of taxonomically reduced taxon set using posterior appendicular characters (90–138, 144–148, and 157). Abbreviations: Aca, *Acanthostega*; Arc, *Archeria*; Asa, *Asaphestera*; Bal, *Balanerpeton*; Bap, *Baphetes*; Bar, *Barameda*; Cae, *Caerorhachis*; Cat, *Catskill humerus*; Cra, *Crassigyrinus*; Den, *Dendrerpeton*; Eoh, *Eoherpeton*; Euc, *Eucritta*; Eus, *Eusthenopteron*; Gly, *Glyptolepis*; Gog, *Gogonasus*; Gre, *Greererpeton*; Hyn, *Hynierpeton*; Ich, *Ichthyostega*; Mar, *Marsdenichthys*; Med, *Medoevia*; Meg, *Megalichthys*; Ony, *Onychodus*; Oss, *Ossinodus*; Pan, *Panderichthys*; Ped, *Pederpes*; Pho, *Pholiderpeton*; Pro, *Proterogyrinus*; Pty, *Ptyonius*; Rhi, *Rhizodus*; Rhp, *Rhizodopsis*; Sap, *Sauropleura*; Sau, *Sauripterus*; Sax, *Saxonerpeton*; Sci, *Scincosaurus*; Sil, *Silvanerpeton*; Ste, *Sterropterygion*; Str, *Strepsodus*; Tik, *Tiktaalik*; Tri, *Trimerorhachis*; Tud, *Tuditanus*; Tul, *Tulerpeton*; Uro, *Urocordylus*; Wes, *Westlothiana*; Wha, *Whatcheeria*.

TEXT-FIG. 2. Results of character partition experiments. Arrows point to the reference value of various indices used in the experiments of character partition from the original matrix (99 pectoral and 55 pelvic characters). A, frequency distribution of incongruent length difference (ILD) values for data sets randomly partitioned 500 times in the same proportions as the original data; B, frequency distribution of Robinson-Foulds (RF) values for data sets randomly partitioned 500 times in the same proportions as the original data; C, frequency distribution of maximum agreement subtree distance (d1) values for data sets randomly

partitioned 500 times in the same proportions as the original data (see text for explanations of the ILD, RF, and d1 indices).

TEXT-FIG. 3. Rarefaction profiles for mean values of four disparity indices and associated 90 percent confidence intervals, based on full character matrix. A, sum of ranges; B, sum of variances; C, root-product of ranges; D, root-product of variances.

TEXT-FIG. 4. Rarefaction profiles for mean values of four disparity indices and associated 90 percent confidence intervals, based on anterior (A–D) and posterior (E, F) appendicular characters. The occurrence of zero-value ranges and variances for scores on some PCo axes made it impossible to calculate product-based indices for the posterior appendicular data. A, E, sum of ranges; B, F, sum of variances; C, root-product of ranges; D, root-product of variances.

TEXT-FIG. 5. Two-dimensional morphospace plots based on pairwise combinations of the first three PCo axes (A–C) and on the first two NMDS axes (D). Plots were based on the whole character set. Black circles represent fish; white circles represent tetrapods. [Abbreviations as in Figure 1.](#)

TEXT-FIG. 6. A, C, density and perspective plot for taxa plotted on the plane delimited by the first two PCo axes; B, D, density and perspective plot for taxa plotted on the plane delimited by the first two NMDS axes. Plots were based on the whole character set. Circles represent fish; triangles represent tetrapods.

TEXT-FIG. 7. Ripley's K function (solid black line) for taxon distributions in the two-dimensional planes delimited by pairwise combinations of the first three PCo axes (PCo1-PCo2, A–C;

PCo1-PCo3, D–F; PCo2-PCo3, G–I) for the full taxon sets (A, D, G), fish only (B, E, H), and tetrapods only (C, F, I). In all plots, the dashed line is the theoretical Poisson distribution of a set of points of identical number to the taxa, and the dotted lines are the upper and lower boundaries of a 95 percent confidence envelope.

FIG. 8. Ripley's K function (solid black line) for taxon distributions in the two-dimensional plane delimited by the first two NMDS axes for the full taxon sets (A), fish only (B), and tetrapods only (C). In all plots, the dashed line is the theoretical Poisson distribution of a set of points of identical number to the taxa, and the dashed lines are the upper and lower boundaries of a 95 percent confidence envelope.

FIG. 9. Goodness-of-fit tests applied to the distribution of unique states in the Ruta (2011) matrix, using maximum likelihood (A) and chi-square optimization (B) criteria. In each graph, the histograms are square root-transformed values of the number of instances of unique states assigned to each individual taxon; the dots show expected counts from a Poisson distribution; the histograms are displaced vertically to match the values of the expected counts, and the amount of displacement is taken to represent the departure from a Poisson distribution.

Formatted: Font: Courier

Formatted: Left, Level 1, Line spacing: Double